干旱区地理
ARID LAND GEOGRAPHY

哈萨克斯坦北部小麦遥感估产方法研究

尹瀚民[1,2]，古丽·加帕尔[1,2,3]，于 涛[1,2]，Jeanine UMUHOZA[1,2]，李 旭[1,2]

（1. 中国科学院新疆生态与地理研究所荒漠与绿洲国家重点实验室，新疆 乌鲁木齐 830011；2. 中国科学院大学，北京 100049；3. 中国科学院中亚生态与环境研究中心，新疆 乌鲁木齐 830011）

摘 要：以哈萨克斯坦北部雨养耕作区为研究靶区，基于春小麦产量统计数据和遥感光谱指数，开展了春小麦估产最优预测时期及植被指数分析，采用回归分析、随机森林、支持向量机及双向循环神经网络模型估算春小麦产量，并对比分析了不同模型的模拟精度。结果表明：北哈萨克斯坦州、阿克莫拉州和库斯塔纳州2007—2016年春小麦估产的最佳预测时期为6月26日—8月5日，该时期是春小麦产量形成的关键时期。北哈萨克斯坦州春小麦估产最优植被指数为7月12日的绿度叶绿素指数（Green chlorophyll index, CI_{green}），阿克莫拉州春小麦估产最优植被指数为8月5日的绿度动态宽波段指数（Green wide dynamic range vegetation index, $WDRVI_{green}$），库斯塔纳州春小麦最优估产植被指数为7月12日的$WDRVI_{green}$。对比分析4种模型模拟春小麦产量的精度，在样本点较少的情况下，双向循环神经网络模型相比其他模型在估算哈萨克斯坦北部三州春小麦产量上精度较高；春小麦产量与植被净初级生产力NPP相关性分析结果显示，北哈萨克斯坦州、阿克莫拉州和库斯塔纳州决定系数R^2在0.50以上面积占比分别为44%、94%和77%，表明上述估产模型可应用于哈萨克斯坦北部三州春小麦估产，尤其是阿克莫拉州和库斯塔纳州。

关 键 词：雨养小麦耕作区；遥感估产；植被指数；回归模型；机器学习；哈萨克斯坦北部

文章编号：

哈萨克斯坦的面粉出口量位于世界第一，被称作中亚粮仓，其北部地区的北哈萨克斯坦州、阿克莫拉州和库斯塔纳州是世界重要的小麦和面粉输出地，小麦的种植结构比例达到86%以上，2010年以来，其小麦和大麦产量居世界第12位，出口量居世界第5位[1]。在中亚的其他国家，由于经济作物占比较大，种植结构不合理且耕地面积有限，严重限制了各国的粮食生产能力，每年都需要向哈萨克斯坦进口大量小麦，才能满足国家粮食需求[1]。近年来，中亚各国的人口增长速度与粮食产量增产速度严重失衡，作为发展中国家，中亚各国经济发展水平本身较低，难以应付国际粮价上涨和国家本身购买力低下带来的一系列社会问题[2]。及时准确估算哈萨克斯坦北部春小麦产量，可为当地政府部门春小麦生产管理提供重要的参考价值，对维护区域粮食安全亦起着不可忽视的作用。

作物估产的方法可以归纳为传统作物估产和遥感作物估产两大类[3]。传统作物估产，一般采取区域人工调查手段，结合农学、气象学，采用统计学方法建立作物估产模式，该方法费时、费力且不利于动态时空监测[3]。现代遥感技术的发展，为区域粮食估产及动态监测提供了有效的手段[4-7]。遥感作物估产方法主要包括经验模型[8-10]、机器学习[11-12]和机理模型[13-14]三大类，其中，经验模型法是利用遥感卫星对作物冠层反射的电磁波信息，计算表征植被特征的多个植被指数，通过与作物的实际产量建立经验模型，寻找估算作物产量的最优植被指数，从而提高作物估产精度。例如Bolton等[15]利用MODIS产品计算2波段增强型植被指数（Enhanced vegetation index 2, EVI2）和归一化水体指数（Normalized difference water index, NDWI），并利用该指数估算了美国中部玉米和大豆产量，发现在玉米返青后

收稿日期：2021-05-20； **修订日期**：2021-08-03
基金项目：中国科学院A类战略性先导科技专项（XDA19030301）资助
作者简介：尹瀚民（1994-），男，硕士研究生，主要从事干旱区农业遥感分析. E-mail: yinhanmin18@163.com
通讯作者：古丽·加帕尔（1973-），女，研究员，主要从事干旱区荒漠植被参量遥感定量反演. E-mail: glmr@ms.xjb.ac.cn

的65~75 d和大豆返青后的80 d左右，植被指数与作物产量相关性最高。Leroux等[16]以非洲萨赫勒地区为研究区，利用MODIS归一化植被指数（Normalized difference vegetation index, NDVI）、地表温度（Land surface temperature, LST）、作物模型SARRA-H模拟的地表生物量和收获指数建立经验统计模型估算作物产量，结果发现组合后的指数在估算产量上优于单独基于NDVI模型。郭锐[17]利用10 a的MOD09A1和MOD16A2产品计算了增强型植被指数（Enhanced vegetation index, EVI）和作物水分胁迫指数（Crop water stress index, CWSI）估算了山东省和各市级尺度的冬小麦产量，其估算精度均不低于89.41%。此外，近些年来，机器学习模型在遥感作物估产方面得到广泛应用，其主要包括随机森林、支持向量机、神经网络等。安秦[18]在长春地区结合HJ-1A/B和Landsat 8 2种卫星影像数据计算了4种植被指数，通过对比曲线统计、多元逐步回归、偏最小二乘回归和神经网络模型，发现神经网络建立的估产模型在稳定性和精度性方面更高。曾妍等[19]利用MODIS产品计算了条件温度植被指数（Vegetation temperature condition index, VTCI）和叶面积指数（Leaf area index, LAI），通过支持向量回归模型估算了关中平原冬小麦产量，模型决定系数达0.88。机理模型作物估产是依据作物的生理生长特征，考虑作物的光合和呼吸作用，并考虑作物生长的环境，例如气温、降水、土壤肥力等因素，模拟作物的生长过程，同时将遥感数据与作物模型有效结合，从而实现对作物产量的预测，例如WOFOST、APSIM、DSSAT、CASA模型等。例如黄健熙等[20]对比了LAI和蒸散（Evapotranspiration, ET）在同化前和同化后融合到SWAP作物模型中进行估算冬小麦产量，发现同化后的精度比同化前高出31%左右。同时马鸿元等[21]还以WOFOST为动态模型，采用集合卡尔曼滤波同化算法，研发了一种遥感LAI与作物模型同化的区域冬小麦产量系统。尽管机理模型虽然功能强大，但所需参数较多，模型计算量大，对计算机能力要求高，本地参数耗时耗力，并且依赖于站点数据，因此在无站点数据区域难以推广[18]。

哈萨克斯坦北部地区缺乏气象观测站点数据和详细的土壤数据，因此难以利用作物生长模型进行春小麦产量估测。经验模型应用较为广泛，其主要是通过建立单一植被指数或者多植被指数的组合与作物实际产量构建线性或者非线性方程，比较其相关性，最终筛选出最佳的产量预测模型，但这种常规模型不一定能满足估测小麦产量的精度要求。机器学习相比常规模型，有其独特优势，例如善于处理高维变量和对变量间的线性和非线性关系有着极强的学习能力，以广泛应用于地理学的研究当中[22]。尽管部分小麦估产的研究中涉及了机器学习模型，但往往仅是一种机器学习模型，很少研究将多种机器学习模型与常规方法进行对比研究，且在哈萨克斯坦北部地区很少有关于春小麦产量的研究报道。本研究基于MODIS遥感数据产品和作物产量统计数据，采用经验模型及机器学习方法，利用线性回归、随机森林、神经网络和支持向量机模型，对比分析了10种植被指数估测春小麦产量的能力，评估春小麦最佳估产时间、最优植被指数及估产模型，以期为当地春小麦生产提供管理依据。

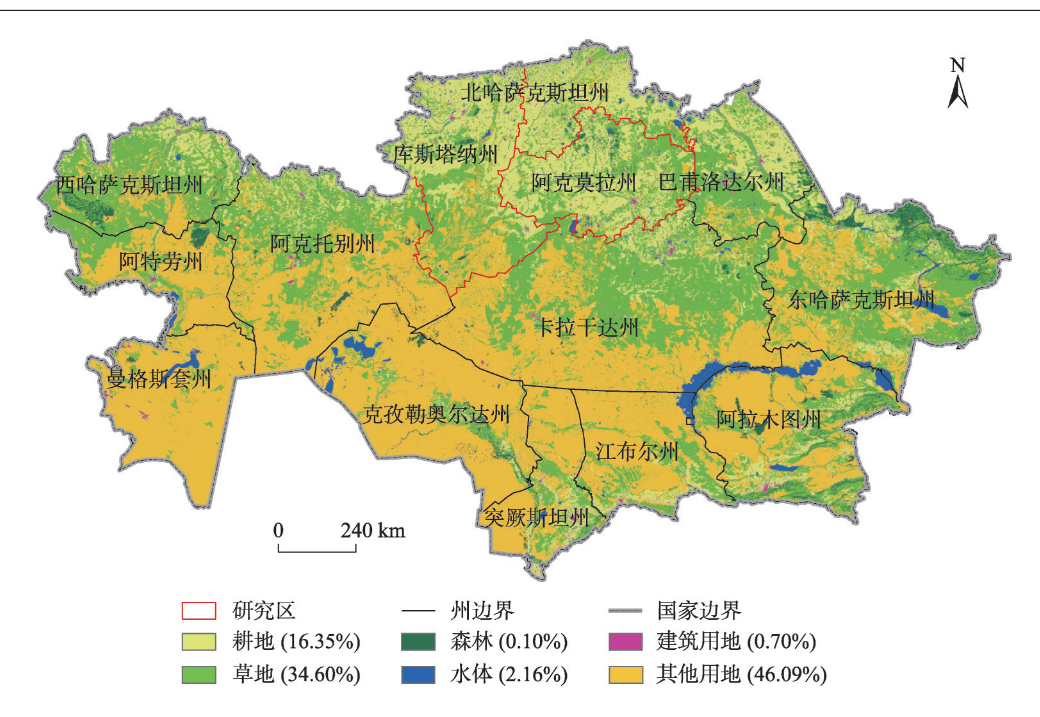
1　研究区概况

哈萨克斯坦北部区域包括北哈萨克斯坦州、库斯塔纳州和阿克莫拉州（图1），位置介于60°03′~74°16′E，48°09′~55°25′N[1]。该区域为温带大陆性气候，夏季降水集中，冬季寒冷漫长[23]。冬季的平均气温在-15~ -20 ℃，最低可达-30 ℃，夏季平均气温18~25 ℃，最高可达30 ℃。北哈萨克斯坦州年平均降水量为407 mm，库斯塔纳州年平均降水量为353 mm，阿克莫拉州年平均降水量为381 mm。伊希姆河和托博尔河流经哈萨克斯坦北部。北哈萨克斯坦州有众多湖泊，都为淡水湖。主要的植被类型为草地、耕地、建筑用地、森林、裸地和水体，分别约占总面积的36%、50%、1%、1%、10%和2%。该地区土壤以黑色土、棕色土为主，土壤肥沃、地势平坦，是世界上重要的小麦出口输出地，由于该地区农田为雨养田，受降水影响大，导致产量极不稳定[24-25]。

2　数据与方法

2.1　数据来源与处理

哈萨克斯坦北部耕作区矢量数据由中国科学院地球大数据科学工程（CASEarth）专项提供（http://data.casearth.cn/），数据生产者为中科院新疆生态与地理研究所遥感研究团队，该数据集以Landsat遥感



注：该图基于自然资源部地图、省级自然资源厅技术审查中心下载的标准地图审图号为GS(2016)2948号制作（下载网址：http://zwfw.ch.mnr.gov.cn/index），底图边界无修改。下同。

图1　哈萨克斯坦北部地理位置及土地利用类型

Fig. 1　Location and types of land use in northern Kazakhstan

影像为数据源，采用面向对象的遥感分类方法，应用影像分割、决策树分类、变化监测等关键技术解译获得，数据精度达到93%以上[26]。计算植被指数所使用的遥感产品是2007—2016年每年6—8月每8 d的MOD09A1和MOD15A2H数据集，其分辨率为500 m。小麦估产采用2007—2016年的植被净初级生产力MOD17A3HGF产品进行间接验证[27]。利用Google Earth Engine（GEE）平台，对MOD09A1、MOD15A2H和MOD17A3HGF产品进行预处理，提取优质像元，以减少云污染对数据质量的影响。

2.2　植被指数计算

选用在作物估产和估算生物量方面应用频率较高的10种植被指数进行最优植被指数分析。哈萨克斯坦北部春小麦的播种日期自5月初开始，8月底到9月初进行春小麦收割。因此，利用GEE平台，下载了春小麦关键生长期6月2日—8月30日每8 d的MOD15A2H叶面积指数产品，并利用MOD09A1产品计算了在此日期内每8 d的其余9种植被指数，对哈萨克斯坦北部三州春小麦进行动态监测。9种植被指数计算公式如表1所示。

2.3　确定最优植被指数和最佳预测时间

通过利用哈萨克斯坦北部三州春小麦实际产量与哈萨克斯坦北部三州耕作区植被指数平均值进行线性拟合，利用决定系数（R^2）和均方根误差（RMSE）作为选择最优植被指数和最佳预测时间的依据。

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-y_i\right)^2}{N}} \tag{1}$$

式中：x_i、y_i分别为冬小麦单产的预测值和实际值（$\mathrm{kg\cdot hm^{-2}}$）；N为年数。

2.4　春小麦估产模型

基于最佳的植被指数和预测时间，分别使用回归、随机森林、支持向量机和双向循环神经网络模型对哈萨克斯坦北部三州进行春小麦产量估测[38]。由于本文所使用的样本量较少，因此使用Bootstrap抽样方法，Bootstrap是一种有放回的均匀抽样，它将数据集中的N个样本数据，有放回的随机抽取N个样本形成新的训练集，利用未抽到的样本作为验证数据集[39]。基于本文的样本数量，数据集循环抽样50次进行模型训练，优化机器学习下的各重要参数。各州选出的最佳植被指数和预测时间作为模型输入，模型输出为春小麦产量。由于本文所使用的是单变量，所以在随机森林模型中不对最佳变量个数

表1 植被指数公式

Tab. 1 Vegetation index formulas

植被指数	公式	参考文献
NDVI	$NDVI=(\rho_2-\rho_1)/(\rho_2+\rho_1)$	[28–30]
EVI2	$EVI2=[2.5\times(\rho_2-\rho_1)]/(\rho_2+2.4\times\rho_1+1)$	[31]
WDRVI	$WDRVI=(\alpha\times\rho_2-\rho_1)/(\alpha\times\rho_2+\rho_1)$，$\alpha$ 取值为0.1	[32]
SANDVI	$SANDVI=\begin{cases}NDVI, & NDVI\leq0.78\\ 0.03\times RVI+0.5363, & NDVI>0.78\end{cases}$ $RVI=\rho_2/\rho_1$	[33]
$WDRVI_{green}$	$WDRVI_{green}=[(\alpha\times\rho_2-\rho_4)/(\alpha\times\rho_2+\rho_4)]+[(1-\alpha)/(1+\alpha)]$，$\alpha$取值为0.3	[34]
CI_{green}	$CI_{green}=(\rho_2/\rho_4)-1$	[35]
DVI	$DVI=\rho_2-\rho_1$	[36]
OSAVI	$OSAVI=(1+1.16)\times[(\rho_2-\rho_1)/(\rho_2+\rho_1+0.16)]$	[37]
NMDI	$NMDI=[\rho_2-(\rho_6-\rho_7)]/[\rho_2+(\rho_6-\rho_7)]$	[8]

注：NDVI为归一化植被指数；EVI2为2波段增强型植被指数；WDRVI为宽动态范围植被指数；SANDVI为饱和校正NDVI指数；$WDRVI_{green}$为绿度动态宽波段指数；CI_{green}为绿度叶绿素指数；DVI为差值植被指数；OSAVI为优化土壤校正植被指数；NMDI为归一化多波段干旱指数；ρ_1、ρ_2、ρ_4、ρ_6和ρ_7分别为MOD09A1产品中的红波段、近红外波段、绿波段、短波红外1和短波红外2波段；α为校正系数；RVI为比值植被指数。下同。

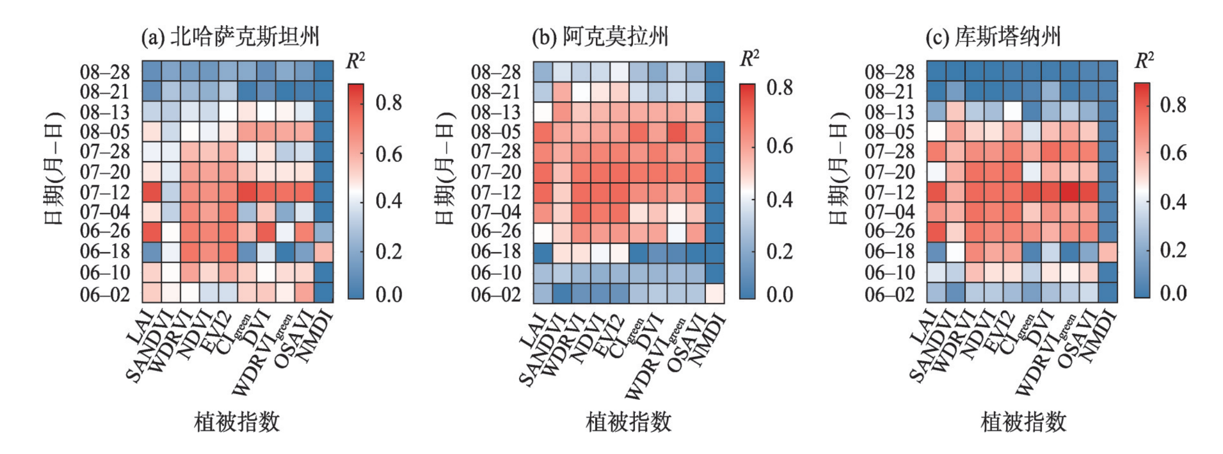
进行训练，而对回归树的个数进行训练。支持向量机的核函数选择线性函数，主要对其代价函数进行训练。双向循环神经网络主要调整神经元个数，利用验证数据集的RMSE决定最佳的模型参数，RMSE越小，模型估算的春小麦产量精度越高。

3 结果与分析

3.1 哈萨克斯坦北部春小麦估产最优植被指数及最佳估产日期

基于多波段的MOD09A1和叶面积指数MOD16A2产品每8 d的光谱信息，计算哈萨克斯坦北部三州耕作区植被指数，利用一元线性回归模型拟合作物产量，期望找出最优植被指数和最佳估产日期。

图2a和表2显示，归一化多波段干旱指数（Normalized multi-band drought index, NMDI）与春小麦产量在三州的拟合效果表现较差。2007—2016年6月2日—8月30日，北哈萨克斯坦州各植被指数与春小麦产量R^2表现为先升高后降低，最优植被指数为7月12日左右的CI_{green}，R^2系数达到0.83，其P值小于



注：LAI为叶面积指数；SANDVI为饱和校正NDVI指数；WDRVI为宽动态范围植被指数；NDVI为归一化植被指数；EVI2为2波段增强型植被指数；CI_{green}为绿度叶绿素指数；DVI为差值植被指数；$WDRVI_{green}$为绿度动态宽波段指数；OSAVI为优化土壤校正植被指数；NMDI为归一化多波段干旱指数；R^2为决定系数。

图2 哈萨克斯坦北部植被指数和春小麦单产决定系数

Fig. 2 Coefficient of determination between the vegetation index and the yield per unit of spring wheat in northern Kazakhstan

表2 哈萨克斯坦北部最优植被指数及最佳估产日期

Tab. 2 Optimal vegetation index and best estimated date of spring wheat yield in northern Kazakhstan

地名	最优植被指数(排名前5位)	最佳观测日期(月-日)	R^2	RMSE/kg
北哈萨克斯坦州	CI_{green}	07-12	0.83	131.8
	LAI	07-12	0.82	135.5
	LAI	06-26	0.80	143.0
	DVI	07-12	0.77	153.6
	OSAVI	07-12	0.76	155.3
阿克莫拉州	$WDRVI_{green}$	08-05	0.80	137.5
	EVI2	07-12	0.78	143.9
	LAI	07-12	0.77	145.0
	WDRVI	07-12	0.76	147.6
	EVI2	07-20	0.76	148.2
库斯塔纳州	$WDRVI_{green}$	07-12	0.88	123.6
	OSAVI	07-12	0.83	150.7
	CI_{green}	07-12	0.81	157.5
	DVI	07-12	0.81	158.2
	LAI	06-26	0.80	163.7

注:R^2为决定系数;RMSE为均方根误差。

0.001,RMSE为131.8 kg。其次是6月26日和7月12日的LAI,与春小麦产量R^2都大于或等于0.80,P值都小于0.001,RMSE分别是143.0 kg和135.5 kg。在北哈萨克斯坦州,春小麦最早可以利用6月26日的LAI指数进行春小麦估产。

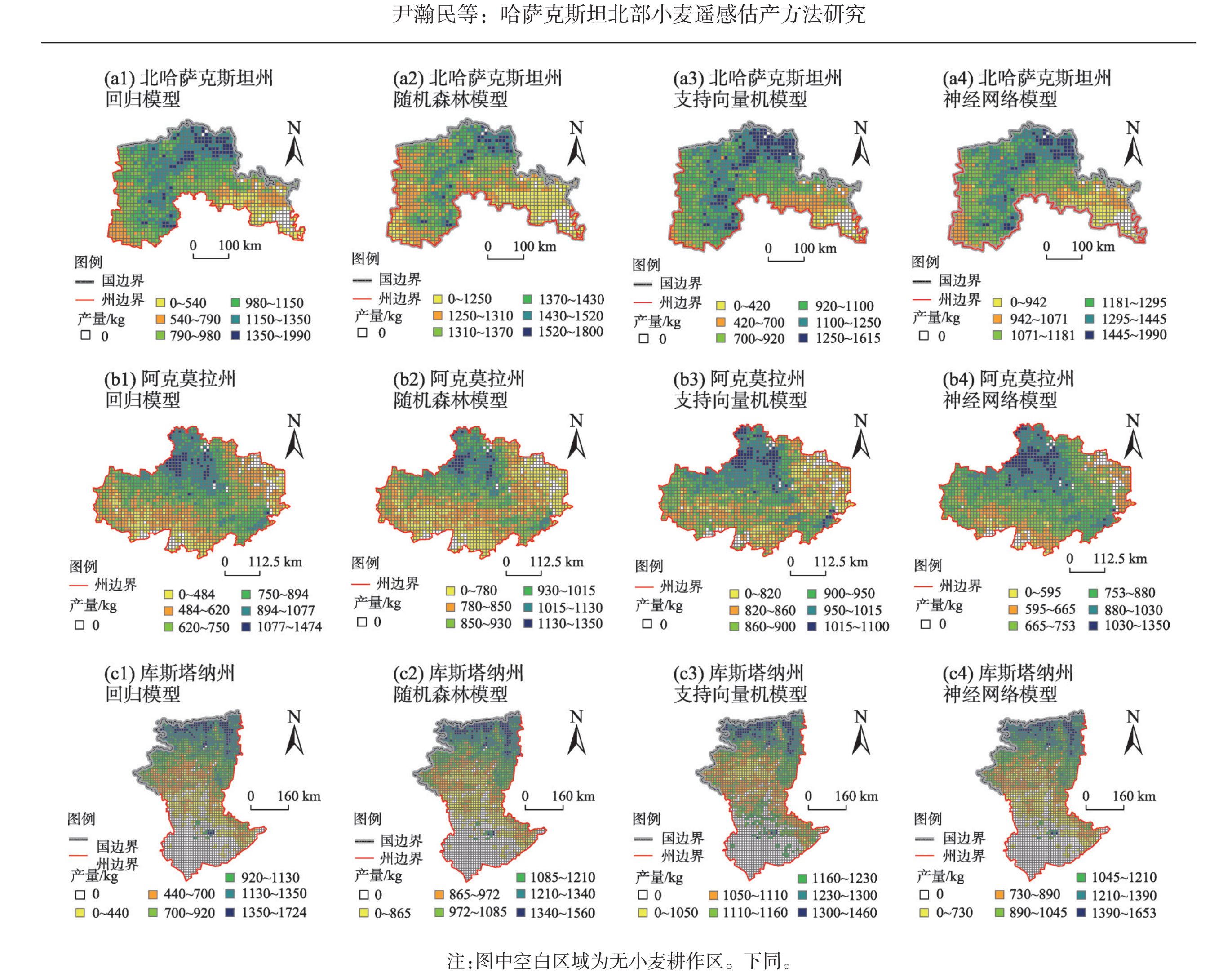
图2b和表2结果表明,除NMDI以外,6月26日以后其他植被指数与春小麦产量R^2在阿克莫拉州开始迅速增加,8月21日之后R^2系数普遍下降,总体来看,7月4日—8月5日植被指数与春小麦产量相关性较好,而6月18日以前各指数与春小麦产量相关性不佳。与春小麦单产拟合最好的植被指数是$WDRVI_{green}$,最佳预测时间在8月5日前后,R^2为0.80,P值小于0.001,RMSE为137.5 kg。此外,在阿克莫拉州,春小麦估产最早可以在7月12日左右进行,宽动态范围植被指数(Wide dynamic range vegetation index, WDRVI)、LAI和EVI2,R^2都超过0.75,RMSE分别为147.6 kg、145.0 kg和143.9 kg。

从图2c和表2表明,库斯塔纳州在6月10日以前和8月13日以后,各指数对春小麦产量拟合性较差,敏感性低。6月26日—7月28日为春小麦产量预测最佳时期,7月12日各指数对春小麦产量的敏感性达到最高,大部分指数P值小于0.001。其中,$WDRVI_{green}$与春小麦产量在7月12日相关性较强,R^2为0.88,P值小于0.001,RMSE为123.6 kg。CI_{green}、差值植被指数(Difference vegetation index, DVI)和优化土壤校正植被指数(Optimized soil adjusted vegetation index, OSAVI)与春小麦产量R^2在7月12日都大于或等于0.81,P值小于0.001,RMSE分别为157.5 kg、158.2 kg和150.7 kg。在库斯塔纳州,春小麦最早可以在6月26日前后利用LAI指数进行春小麦估产。

经对比分析,北哈萨克斯坦州春小麦产量预测最佳日期为7月12日,最优植被指数为CI_{green},阿克莫拉州春小麦产量预测最佳日期为8月5日,最优植被指数为$WDRVI_{green}$,库斯塔纳州春小麦产量预测最佳日期为7月12日,最优植被指数为$WDRVI_{green}$。

3.2 哈萨克斯坦北部耕作区春小麦估产最优模型及空间分布特征

利用线性回归、随机森林、支持向量机和神经网络模型对哈萨克斯坦北部三州耕作区进行春小麦产量估测,以期找出最佳的产量预测模型。图3采用ArcGIS自然间断点进行分类,结果表明,在北哈萨克斯坦州利用线性回归、支持向量机和神经网络模型模拟出的产量在空间分布特征上相似性较高,随机森林模型模拟的春小麦高产区分布较为离散,聚集性不强,在西部和东南部低产区分布较多。总体而言,4种模型模拟的春小麦产量在东南部地区相对较低,中北部地区较高。在阿克莫拉州,支持向量机模型在南部模拟的产量分布上出现



注：图中空白区域为无小麦耕作区。下同。

图3　2007—2016年哈萨克斯坦北部多年平均估算春小麦产量

Fig. 3　Estimated annual average production of spring wheat yield in northern Kazakhstan from 2007 to 2016

破碎化特征，并且支持向量机模型估算的产量区域差异较小。除支持向量机模型外，其他3种模型模拟的春小麦产量在阿克莫拉州的北部相较于其他区域产量较高，东部和南部地区春小麦产量较低，回归模型和随机森林模型模拟的产量在空间分布上相似性较高。在库斯塔纳州，4种模型估算的春小麦产量空间分布特征具有较高的相似性且产量表现为北部较高，南部较低。同样支持向量机模型模拟的产值区域性差异较小。综合上述分析，尽管模型之间有所差别，但模拟的春小麦产量在空间分布特征上具有一致性，通过对比发现，随机森林模型和支持向量机模型模拟的春小麦产量在三州有不同程度的缺陷，回归模型和神经网络模型在估算哈萨克斯坦北部三州春小麦产量上有较好的表现。

3.3　春小麦估产精度分析

植被净初级生产力NPP通常乘以作物收获指数来估算作物产量[27]。由于本研究春小麦产量统计数据有限，且主要用于模型输入参数，故为评价春小麦估产模型的优劣，研究分析不同模型模拟春小麦产量与NPP（MOD17A3HGF）之间的相关性，期望找出最佳产量估测模型（图4），并对各州NPP与春小麦产量相关性统计结果进行分级（图5），具体分为微弱相关（0.0~0.3）、低度相关（0.3~0.5）、显著相关（0.5~0.8）和高度相关（0.8~1.0）。

研究表明，在北哈萨克斯坦州，回归模型、支持向量机模型和神经网络模型相比随机森林模型能够更好的预测春小麦产量，其中表现最差的是随机森林模型，低度相关性区域占北哈萨克斯坦州耕作区近68%，显著相关区域占比仅8%。显著相关区主要分布在北哈萨克斯坦州西部和东南部地区，而表现为低度相关和微弱相关区域主要分布在北哈萨克斯坦州的北部。

图4 哈萨克斯坦北部估算春小麦产量空间验证

Fig. 4 Spatial validation of spring wheat yield estimation in northern Kazakhstan

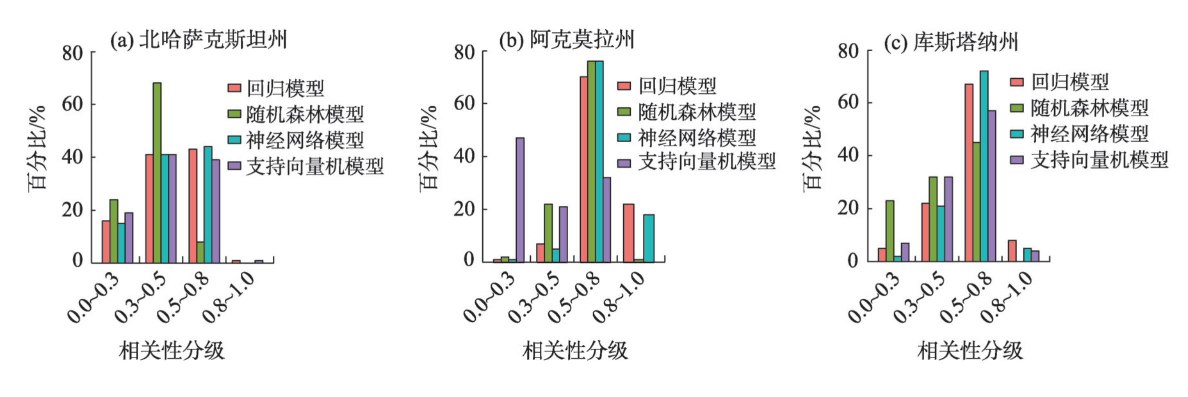


图5 NPP与春小麦产量相关性分级

Fig. 5 Classification of the correlation between NPP and spring wheat yield

除支持向量机模型外，其余3种模型在阿克莫拉州都得到了很好的验证结果，尤其是回归模型和神经网络模型，大部分地区NPP与估算的春小麦产量相关性都在0.5以上。表现最佳的是神经网络模型，估测的产量与NPP显著相关和高度相关占比总和最高为94%，其次是回归模型为92%，再次是随机

森林模型为77%。从空间分布来看,无论是高产区还是低产区,回归模型、神经网络模型和随机森林模型预测精度都较好,对于了解该地区高产和低产田的分布位置和产量,进而合理安排该地区的农业生产活动具有重要的参考价值。

在库斯塔纳州,回归模型、支持向量机模型和神经网络模型模拟的精度较高,随机森林模型模拟的结果不理想,其微弱相关占比为23%,主要分布在库斯塔纳州南部地区,而其他模型在库斯塔纳州南部地区表现相对较好,大部分区域相关性都在0.50以上。而在库斯塔纳州的北部地区,4种模型估测的春小麦产量与NPP相关性表现为低度相关,而根据春小麦产量统计结果来看,北部区域实际为库斯塔纳州的高产区,该种现象与北哈萨克斯坦州的北部估测产量验证精度较低具有一致性。

4 讨 论

遥感作物估产主要是利用遥感影像建立表征作物因子与产量之间的模型体系,从而实现对作物产量预测,可使用单植被指数或者多植被指数组合,亦可与影响春小麦产量相关的气温、降水和土壤肥力等因子估算春小麦产量[40],但使用多因子估算春小麦产量不仅对样本数据量有要求,同时也要考虑多因子之间的共线性问题,在样本量较少的情况下,使用多因子组合不仅不会提高估产精度,而且会产生较大偏差。

通过利用机器学习模型和回归模型估算春小麦产量发现,各模型估算的春小麦高产区和低产区具有极大的相似性,4种模型在北哈萨克斯坦州估测产量的能力普遍较差,尤其是在北哈萨克斯坦州的北部地区,该地区是春小麦高产区,同样的现象也发生在库斯塔纳州的北部,经过对降雨数据分析以及地面调查结果发现,该地区雨量较大,对MODIS产品数据质量造成了影响,也有可能是由于验证数据MOD17A3HGF产品在该区域出现误差,还有一个重要原因是由于北哈萨克斯坦州小麦的种植结构比例约在86%,而库斯塔纳州和阿克莫拉州小麦种植比例在93%以上,在北哈萨克斯坦州有大面积的大麦、荞麦和土豆种植,不能完全视为春小麦种植区,从而影响估产模型的精度。

从模型角度来分析,回归随机森林模型无法实现训练数据集范围外的预测,且由于本文数据样本量较小,在进行随机采样构建回归树时,会出现大量相似的回归树,并且高产和低产的样本有时不会被随机抽到,所以在估测产量时低产或高产的区域相比其他模型偏大或者偏小。支持向量机模型在寻求结构化风险最小的方式来矫正模型[41],并且最终决策函数只由少数的支持向量所决定,所以有可能将几个中度产值的样本数据建立超平面估算作物产量,导致估算的作物产量空间差异性较小。本文发现,在样本点较少的情况下,双向循环神经网络和回归模型更加可靠,回归模型不仅可以节省大量机器学习进行模型参数优化的时间,并且在模型的强健性上不输于机器学习模型,但回归模型的可移植性不高,从模型验证的整体情况来看,机器学习中的神经网络模型估算精度略优于其他3种模型。由于缺乏地面实测产量对其进行验证,NPP作为表征地上生物量的重要指标,尽管MOD17A3HGF产品在某些区域存在精度限制,但在地面实测数据缺乏的条件下,可以作为验证春小麦产量的重要数据源。

5 结 论

(1) 除NMDI指数外,植被指数与春小麦单产相关性较强的时间段主要集中在6月26日—8月5日之间,主要是因为该阶段哈萨克斯坦北部春小麦进入抽穗扬花到逐渐成熟时期,是春小麦产量形成的关键时期。春小麦估产最优植被指数分别是北哈萨克斯坦州7月12日CI_{green},阿克莫拉州为8月5日的$WDRVI_{green}$,库斯塔纳州为7月12日的$WDRVI_{green}$。

(2) 从4种模型模拟的春小麦产量空间分布可知,模型模拟的产量虽有差距,但高产和低产的分布范围具有一致性,支持向量机模型在模拟产量时有可能造成产量区域差距较小,而随机森林模型容易出现高产区或者低产区偏大或者偏小的特点,回归模型和神经网络模型稳定性更高。

(3) NPP与不同模型估测的春小麦产量进行相关性分析,结果表明,在阿克莫拉州和库斯塔纳州,最佳春小麦估产模型为神经网络模型,其次为回归模型,支持向量机模型和随机森林模型精度较低。而在北哈萨克斯坦州,4种模型的预测精度与其他二州相比相对较差,相对而言,双向循环神经网络

模型表现最佳。

参考文献(References)

[1] 李宁. 哈萨克斯坦小麦产业的发展与问题[J]. 粮食流通技术, 2013(2): 37-42. [Li Ning. Study on Kazakhstan's wheat industry and its problems[J]. Grain Distribution Technology, 2013(2): 37-42.]

[2] 马骏, 龚新蜀. 中亚国家粮食安全问题研究[J]. 世界农业, 2014(8): 22-26. [Ma Jun, Gong Xinshu. Research on food security in Central Asian countries[J]. World Agriculture, 2014(8): 22-26.]

[3] 周清波. 国内外农情遥感现状与发展趋势[J]. 中国农业资源与区划, 2004, 25(5): 12-17. [Zhou Qingbo. Status and tendency for development in remote sensing of agriculture situation[J]. Journal of China Agricultural Resources and Regional Planning, 2004, 25(5): 12-17.]

[4] Anup K P, Lim C, Ramesh P S, et al. Crop yield estimation model for Iowa using remote sensing and surface parameters[J]. International Journal of Applied Earth Observation and Geoinformation, 2006, 8(1): 26-33.

[5] Michael S R. Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR[J]. International Journal of Remote Sensing, 1992, 13(18): 3871-3879.

[6] Michael S R. Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability[J]. International Journal of Remote Sensing, 1997, 18(5): 1059-1077.

[7] 焦险峰, 杨邦杰, 裴志远, 等. 基于植被指数的作物产量监测方法研究[J]. 农业工程学报, 2005, 21(4): 104-108. [Jiao Xianfeng, Yang Bangjie, Pei Zhiyuan, et al. Monitoring crop yield using NOAA/AVHRR based vegetation indices[J]. Transactions of the Chinese Society of Agricultural Engineering, 2005, 21(4): 104-108.]

[8] Tuvdendorj B. Determination of appropriate remote sensing indices for spring wheat yield estimation in Mongolia[J]. Remote Sensing, 2019, 11(21): 2568-2570.

[9] 李军玲, 郭其乐, 彭记永. 基于MODIS数据的河南省冬小麦产量遥感估算模型[J]. 生态环境学报, 2012, 21(10): 1665-1669. [Li Junling, Guo Qile, Peng Jiyong. Remote sensing estimation model of Henan Province winter wheat yield based on MODIS data[J]. Ecology and Environmental Sciences, 2012, 21(10): 1665-1669.]

[10] 谭昌伟. 基于开花期卫星遥感数据的大田小麦估产方法比较[J]. 中国农业科学, 2017, 50(16): 3101-3109. [Tan Changwei. Comparison of the methods for predicting wheat yield based on satellite remote sensing data at anthesis[J]. Scientia Agricultura Sinica, 2017, 50(16): 3101-3109.]

[11] Uno Y. Artificial neural networks to predict corn yield from compact airborne spectrographic imager data[J]. Computers and Electronics in Agriculture, 2005, 47(2): 149-161.

[12] 王恺宁, 王修信. 多植被指数组合的冬小麦遥感估产方法研究[J]. 干旱区资源与环境, 2017, 31(7): 44-49. [Wang Kaining, Wang Xiuxin. Research on winter wheat yield estimtion with the multipy remote sensing vegetation index combination[J]. Journal of Arid Land Resources and Environment, 2017, 31(7): 44-49.]

[13] 徐春萌. 基于DSSAT作物模型的中美大豆主产区单产模拟与验证[J]. 农业工程学报, 2021, 37(3): 132-139. [Xu Chunmeng. Simulations and validations of the soybean yields per unit area using DSSAT crop model in the major soybean producing areas of China and America[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(3): 132-139.]

[14] 赵军. 基于净初级生产力的春小麦生产潜力及估产研究——以甘肃省白银区为例[J]. 干旱地区农业研究, 2015, 33(1): 199-204. [Zhao Jun. Potential productivity and yield estimation of spring wheat based on a net primary production model: Taking Baiyin District in Gansu Province as an example[J]. Agricultural Research in the Arid Areas, 2015, 33(1): 199-204.]

[15] Bolton D K, Friedl M A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics[J]. Agricultural and Forest Meteorology, 2013, 173: 74-84.

[16] Leroux L. Crop monitoring using vegetation and thermal indices for yield estimates: Case study of a rainfed cereal in semi-arid west Africa[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(1): 347-362.

[17] 郭锐. 山东省冬小麦单产监测与预报方法研究[J]. 农业机械学报, 2020, 51(7): 156-163. [Guo Rui. Monitoring and forecasting method of winter wheat yield in Shandong Province[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 156-163.]

[18] 安秦. 长春地区玉米产量遥感估算方法研究[D]. 长春: 吉林大学, 2018. [An Qin. Research on methods of maize yield estimation by remote sensing in Changchun region[D]. Changchun: Jilin University, 2018.]

[19] 曾妍, 王迪, 赵小娟, 等. 基于支持向量回归的关中平原冬小麦估产研究[J]. 中国农业信息, 2019, 31(6): 10-20. [Zeng Yan, Wang Di, Zhao Xiaojuan, et al. Study on yield prediction of winter wheat in Guanzhong Plain based on SVR[J]. China Agricultural Informatics, 2019, 31(6): 10-20.]

[20] 黄健熙, 马鸿元, 田丽燕, 等. 基于时间序列LAI和ET同化的冬小麦遥感估产方法比较[J]. 农业工程学报, 2015, 31(4): 197-203. [Huang Jianxi, Ma Hongyuan, Tian Liyan, et al. Comparison of remote sensing yield estimation methods for winter wheat based on assimilating time-sequence LAI and ET[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(4): 197-203.]

[21] 马鸿元, 黄健熙, 黄海, 等. 基于历史气象资料和WOFOST模型的区域产量集合预报[J]. 农业机械学报, 2018, 49(9): 257-266. [Ma Hongyuan, Huang Jianxi, Huang Hai, et al. Ensemble forecasting of regional yield of winter wheat based on WOFOST model using historical metrological dataset[J]. Transactions of the Chi-

nese Society for Agricultural Machinery, 2018, 49(9): 257-266.]

[22] 杨练兵, 郑宏伟, 罗格平, 等. 基于遗传算法优化BP神经网络的土壤盐渍化反演[J]. 地理与地理信息科学, 2021, 37(2): 21-37. [Yang Lianbing, Zheng Hongwei, Luo Geping, et al. Retrieval of soil salinity content based on BP neural network optimized by genetic algorithm[J]. Geography and Geo-Information Science, 2021, 37(2): 21-37.]

[23] Dara A. Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series[J]. Remote Sensing of Environment, 2018, 213: 49-60.

[24] Guo H. Determining variable weights for an optimal scaled drought condition index (OSDCI): Evaluation in Central Asia[J]. Remote Sensing of Environment, 2019, 231: 1-17.

[25] Yin H. Monitoring fire regimes and assessing their driving factors in Central Asia[J]. Journal of Arid Land, 2021, 13(5): 500-515.

[26] 韩其飞, 罗格平, 白洁, 等. 基于多期数据集的中亚五国土地利用/覆盖变化分析[J]. 干旱区地理, 2012, 35(6): 909-918. [Han Qifei, Luo Geping, Bai Jie, et al. Characteristics of land use and cover change in Central Asia in recent 30 years[J]. Arid Land Geography, 2012, 35(6): 909-918.]

[27] Yao F. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain[J]. Physics and Chemistry of the Earth, 2015, 87: 142-152.

[28] Mkhabela M S, Mashinini N N. Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR[J]. Agricultural and Forest Meteorology, 2005, 129(1): 1-9.

[29] 田艳君, 石莹, 帅艳民, 等. 基于遥感时序特征的地表覆被信息提取[J]. 干旱区地理, 2021, 44(2): 450-459. [Tian Yanjun, Shi Ying, Shuai Yanmin, et al. Land cover information retrieval from temporal features based remote sensing images[J]. Arid Land Geography, 2021, 44(2): 450-459.]

[30] 王华, 杨乾鹏, 田云杰, 等. 基于多时相Landsat影像的中亚地区植被覆盖遥感监测[J]. 干旱区地理, 2020, 43(4): 1023-1032. [Wang Hua, Yang Qianpeng, Tian Yunjie, et al. Vegetation coverage monitoring in the Central Asian countries using multi-temporal Landsat images[J]. Arid Land Geography, 2020, 43(4): 1023-1032.]

[31] Jiang Z. Development of a two-band enhanced vegetation index without a blue band[J]. Remote Sensing of Environment, 2008, 112 (10): 3833-3845.

[32] Gitelson A A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation[J]. Journal of Plant Physiology, 2004, 161(2): 165-173.

[33] Gu Y. NDVI saturation adjustment: A new approach for improving cropland performance estimates in the Greater Platte River Basin, USA[J]. Ecological Indicators, 2013, 30: 1-6.

[34] Gitelson A A. Remote estimation of crop gross primary production with Landsat data[J]. Remote Sensing of Environment, 2012, 121: 404-414.

[35] Gitelson A A. Remote estimation of canopy chlorophyll content in crops[J]. Geophysical Research Letters, 2005, 32(8): 1-4.

[36] Mirchooli F. Spatial distribution dependency of soil organic carbon content to important environmental variables[J]. Ecological Indicators, 2020, 116: 1-5.

[37] Jin Y. Remote sensing-based biomass estimation and its spatio-temporal variations in temperate grassland, northern China[J]. Remote Sensing, 2014, 6(2): 1496-1513.

[38] 高煜堃. 基于机器学习及多源数据的亚热带典型区域森林地上生物量估测研究[D]. 杭州: 浙江农林大学, 2018. [Gao Yukun. Aboveground forest biomass estimation based on machine learning algorithms and multi-source data in a typical subtropical region [D]. Hangzhou: Zhejiang Agriculture and Forestry University, 2018.]

[39] 郜燕芳, 李俊明, 刘东伟, 等. 基于随机森林模型的城市不透水面提取研究——以呼和浩特市为例[J]. 冰川冻土, 2018, 40(4): 828-836. [Gao Yanfang, Li Junming, Liu Dongwei, et al. Research on extraction of urban impervious surface based on random forest model: A case study in Hohhot[J]. Journal of Glaciology and Geocryology, 2018, 40(4): 828-836.]

[40] 徐新刚, 吴炳方, 蒙继华, 等. 农作物单产遥感估算模型研究进展[J]. 农业工程学报, 2008(2): 290-298. [Xu Xingang, Wu Bingfang, Meng Jihua, et al. Research advances in crop yield estimation models based on remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018(2): 290-298.]

[41] 徐逸, 董轩妍, 王俊杰. 4种机器学习模型反演太湖叶绿素a浓度的比较[J]. 水生态学杂志, 2019, 40(4): 48-57. [Xu Yi, Dong Xuanyan, Wang Junjie. Use of remote multispectral imaging to monitor chlorophyll-a in Taihu Lake: A comparison of machine learning models[J]. Journal of Hydroecology, 2019, 40(4): 48-57.]

Wheat yield estimation with remote sensing in northern Kazakhstan

YIN Hanmin[1,2], Guli JIAPAER[1,2,3], YU Tao[1,2], Jeanine UMUHOZA[1,2], LI Xu[1,2]

(1. State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Research Center for Ecology and Environment of Central Asia of Chinese Academy of Sciences, Urumqi 830011, Xinjiang, China)

Abstract: Kazakhstan's flour export volume ranks first in the world and is called the Central Asian granary. Its northern regions, including North Kazakhstan, Aqmola, and Qostanay, are the world's important wheat and flour exporters. The proportion of wheat planting structures has reached 86%. Since 2010, its wheat and barley output has ranked 12th in the world, and its export volume has ranked 5th in the world. However, the region is rain-fed and lacks effective irrigation measures. It often suffers from drought stress due to its location in the monsoon climatic zone, resulting in a large-scale reduction in spring wheat production, which severely restricts the economic development of countries that rely on wheat imports. The estimation of wheat production plays a vital crucial role in promoting regional food security and social stability, especially in responding to the food crisis in the post-epidemic era and achieving zero hunger advocated by the United Nations. In this paper, the rain-fed farming area in northern Kazakhstan is used as the research target area. An analysis of the optimal vegetation index for spring wheat yield estimation was conducted based on the statistical spring wheat yield and vegetation index. Various methods, such as regression, random forest, support vector machine, and neural network, are evaluated for the accuracy of wheat yield estimation. In North Kazakhstan, the best vegetation index for estimating spring wheat yield is the greenness chlorophyll index from 2007 to 2016. The forecast time can be advanced to July 12th from 2007 to 2016. In Aqmola, the best vegetation index for estimating spring wheat yield is the greenness dynamic wide-band vegetation index ($WDRVI_{green}$). The forecast time can be advanced to August 5th. In Qostanay, the best vegetation index for estimating spring wheat yield is the $WDRVI_{green}$. The forecast time can be advanced to July 12th. In this paper, MODIS NPP products are selected to verify the estimation results. Through correlation analysis with NPP shows that the neural network has higher accuracy in estimating spring wheat yield in the three northern states of Kazakhstan than other models. Neural networks accounted for 44%, 94%, and 77% of correlations in North Kazakhstan, Aqmola, and Qostanay states, respectively.

Key words: rain-fed wheat farming area; remote sensing yield estimation; vegetation index; regression model; machine learning; northern Kazakhstan